Volume 194, number 2 FEBS 3240 January 1986

Similarities between a conserved sequence element of homoeo boxes and other genes

Panagiotis A. Tsonis* and John D. Lambris

*La Jolla Cancer Research Foundation, 10901 North Torrey Pines Road, La Jolla, CA 92037 and Scripps Clinic and Research Foundation, Department of Immunology, 10666 North Torrey Pines Road, La Jolla, CA 92037, USA

Received 16 September 1985

We report here the homology of different genes with an 18-nucleotide sequence element derived from a conserved region of the homoeo boxes. Possible evolutionary relationships are discussed.

Homoeo box *Ancestor gene* *IgV_H primordial block*

1. ANALYSIS AND DISCUSSION

The homoeo box is a 180-bp region found first in the homeotic genes of *Drosophila* and subsequently in several other species including human [1–4]. A comparison of all homoeo box sequences reported to date shows that there are 2 conserved regions among all homoeo boxes, the first from 40 to 75 and the second from 120 to 165 nucleotide [5–9]. We have found that an 18-nucleotide sequence element from the first conserved region of the homoeo box is highly conserved in many other genes.

First, we observed that a significant homology exists between the 18-nucleotide sequence element of the homoeo box and regions from recently reported *Drosophila* RNA species detected by a v-myc probe [10]. Both the maternally expressed *Drosophila* RNA and the v-myc [11] corresponding regions show an 83% homology with the 18-nucleotide sequence element from the homoeo boxes. The above observation led us to examine the presence of that element in other genes as well. We used the GeneBank to screen and select sequences showing homology to the conserved 18-nucleotide element of the homoeo box. Only sequences with more than 80% homology were considered.

Several genes were found to contain a region highly homologous to the 18-nucleotide element from the homoeo boxes (fig.1). Interestingly, the rat α_1-acid glycoprotein gene (α_1AGP) contains a region which is 100% homologous [12] to that of the homoeo boxes. Other homologous regions were found in the genes of myosin, calcitonin and cytochrome oxidase (fig.1). In addition, an element from the mouse Ig germline DJC region [13] was found to be 89% homologous to homoeo boxes. The 18-nucleotide element contains identical or very similar repeats to that of a 20-nucleotide sequence (AGCTG) (AGCTG) (AGCTG) (GGGTG), which is considered to be one of the few ultimate ancestors of all euchromatic DNAs [14], and the primordial sequence of intergenic spacers, as judged from theoretical studies employing the immunoglobulin class switch sequence [15]. Moreover, theoretical studies have shown that the primordial building block of the family of immunoglobulin variable region (IgV_H) was the 48-base-long sequence TTC AGC AGC CTG ACT GGA TAT GAC CTG GAG TGG ACT TAC TCG CCA AGA [14]. The conserved 18-nucleotide element shows high homology with the above primordial building block (fig.1), and this suggests that the 18-nucleotide present in all genes mentioned in table 1 is derived from the same ancestor sequence which might be the one that built up the immunoglobulin superfamily. Interestingly, other

Published by Elsevier Science Publishers B.V. (Biomedical Division)

Homoeo box (40-57)	CTG	GAG	CTG	GAG	AAG	GAG
	***	***	***	***	***	***
Rat α_1-acid-glycoprotein (607-624)	CTG	GAG	CTG	GAG	AAG	GAG
	***	***	* *	***	***	**
Rabbit α-myosin heavy chain (7-24) and rabbit β-myosin heavy chain (7-24)	CTG	GAG	CAG	GAG	AAG	AAG
	* *	***	***	***	**	***
Rat cardiac myosin heavy chain (928-945)	CGG	GAG	CTG	GAG	AAT	GAG
	***	***	* *	***	**	**
Rat calcitonin mRNA (190-207)	CTG	GAG	CAG	GAG	GAG	GAA
	**	***	**	***	**	***
Drosophila RNA (71-88)	GTG	GAG	GTG	GAG	CAG	GAG
	***	***	**	* *	* *	***
v-myc (277-294)	CTG	GAG	ATG	GTG	ACG	GAG
	**	***	***	***	***	* *
Mouse Ig switch region (1533-1550)	ATG	GAG	CTG	GAG	AAG	GTG
	***	**	***	***	*	**
IgV_H primodial block	CTG	GAC	CTG	GAG	TG	GAC
		12 22				
	***	***	***	**	**	**
Rat cytochrome oxidase (792-809)	CTG	GAG	CTG	GAA	CAG	GAT

Fig.1.

regions of the α_1AGP molecule exhibit weak sequence homology to immunoglobulin as well as to part of the variable domains of the mouse T-cell receptor and the EGF receptor [16].

The amino acid sequence encoded by the 18-nucleotide element is Leu-Glu-Leu-Glu-Lys-Glu. Although the occurrence of an identical hexapeptide in unrelated proteins is significant [17], one should be very careful in postulating evolutionary relationships and functional similarities. With the conserved region as our alignment point, we extended our study to the vicinity of this region, looking for similarities in the amino acid sequence. The sequences are shown in fig.2, and calculations on the percentage homology are given in table 1. Considering the favored amino acid substitutions, we can see that considerable homology exists at the protein level as well.

Also of interest is the Arg-Arg-Arg-Arg sequence found in the homoeo boxes between the conserved regions. Identical or similar quadruplets

Fig.2.

Table 1

Amino acid homology between HHB_1, α_1AGP, rat MHC, v-myc and pcal

	% similarity (identity)			
	α_1AGP	rat MHC	v-myc	pcal
HHB_1	38 (22)	24 (15)	24 (11)	19 (11)
α_1AGP	–	28 (14)	21 (14)	29 (16)
Rat MHC	–	–	22 (6)	35 (19)
v-myc	–	–	–	20 (9)
pcal	–	–	–	–

HHB_1, Human homoeo box [4]; α_1AGP, rat α_1-glycoprotein [12]; MHC, myosin heavy chain [25]; pcal, rat calcitonin [26]

of basic residues have been found in mouse and human C3 and C4, at both the junctions between the β- and α- and between the α- and γ-subunits [18,19]. The same sequence has been also found in blood coagulation factor X [20], human calcitonin [21] and insulin [22]. The above sequence has been found to be a signal for the maturation process of these proteins. This suggests that the tetra-arginine sequence found between the conserved regions of homoeo boxes may play a similar role.

The role of the homoeo box is not known. However, its presence in the homoeotic genes of *Drosophila* has implied that the homoeo box should be connected with cell lineage, differentiation and morphogenesis. Furthermore, transcription of homoeo box-containing genes during retinoic acid-induced differentiation of F9 teratocarcinoma cells also connects the presence of the homoeo box with differentiation in mammalian cells [23]. Nothing is yet known about the function of the protein domain which is encoded by the homoeo box, but again its conservation among all homoeo boxes in different species implies a very specific function (a signal or recognition of some kind), especially when the two highly conserved regions of the homoeo boxes are conserved. The present observations regarding the presence of a conserved element from the homoeo boxes in various genes could imply such a speculation. Finally, this region should be avoided when synthetic oligonucleotides or peptides are to be made. The homology reported here warns for unspecific reactivity of the probes. Sequences from the second conserved region of homoeo boxes (120–141 nucleotides) do not show any significant homology to nucleotide sequences of other genes (not shown).

ACKNOWLEDGEMENTS

We thank Dr P. Singer for his help with the GeneBank screening and Charlotte Gott for typing the manuscript. This work was kindly supported by cooperation of and grants to Dr E.D. Adamson from the NIH (CA 28427 and HD 18782).

REFERENCES

[1] McGinnis, W., Levine, M.S., Hafen, E., Kuroiwa, A. and Gehring, W.J. (1984) Nature 308, 428–433.
[2] Fjose, A., McGinnis, W. and Gehring, W. (1985) Nature 313, 284–289.
[3] McGinnis, W., Garber, R.L., Wirz, J., Kuroiwa, A. and Gehring, W.J. (1984) Cell 37, 403–408.
[4] Levine, M., Rubin, G.M. and Tjian, R. (1984) Cell 38, 667–673.
[5] McGinnis, W., Hart, C.P., Gehring, W.J. and Ruddle, F.H. (1984) Cell 38, 675–680.
[6] Laughon, A. and Scott, M.P. (1984) Nature 310, 25–31.
[7] Poole, S.J., Kauvar, L.M., Drees, B. and Kornberg, T. (1985) Cell 40, 37–43.
[8] Carrasco, A.E., McGinnis, W., Gehring, W.J. and DeRobertis, E.M. (1984) Cell 37, 409–414.
[9] Muller, M.M., Carrasco, A.E. and DeRobertis, E.M. (1984) Cell 39, 157–162.
[10] Madhavan, K., Bilodeau-Wentworth, D. and Wadsworth, S.C. (1985) Mol. Cell Biol. 5, 7–16.
[11] Alitalo, K., Bishop, J.M., Smith, D.H., Chen, E.Y., Colby, W.W. and Levinson, A.D. (1983) Proc. Natl. Acad. Sci. USA 80, 100–104.
[12] Ricca, G.A. and Taylor, J.M. (1981) J. Biol. Chem. 256, 11199–11202.
[13] Kataoka, T., Miyata, T. and Honjo, T. (1981) Cell 23, 357–368.
[14] Ohno, S. (1982) Perspect Biol. Med. 25, 559–572.
[15] Ohno, S. (1981) Differentiation 18, 65–74.
[16] Toh, H., Hayashida, H., Kikuno, R., Yasunaga, T. and Miyata, T. (1985) Nature 314, 199.
[17] Kabsch, W. and Sander, C. (1984) Proc. Natl. Acad. Sci. USA 81, 1075–1078.
[18] Domdey, H., Wiebauer, K., Kazmaier, M., Muller, V., Odink, K. and Fey, G. (1982) Proc. Natl. Acad. Sci. USA 79, 7619–7623.
[19] Whitehead, A.S., Goldberger, G., Woods, D.E., Markham, A.F. and Colten, H.R. (1983) Proc. Natl. Acad. Sci. USA 80, 5387–5391.

[20] Leytus, S.P., Chung, D.W., Kisiel, W., Kurachi, K. and Davie, E.W. (1984) Proc. Natl. Acad. Sci. USA 81, 3699–3702.
[21] Edbrooke, M.R., Parker, D., McVey, J.H., Riley, J.H., Sorenson, G.D., Pettengill, O.S. and Craig, R.K. (1985) EMBO J. 4, 715–724.
[22] Ullrich, A. et al. (1985) Nature 313, 756–761.
[23] Colberg-Poley, A.M., Voss, S.D., Chowdhury, K. and Gruss, P. (1985) Nature 314, 713–718.
[24] Sinha, A.M., Umeda, P.K., Kauinsky, C.J., Rajama-nickam, C., Hsu, H.J., Jakovcic, S. and Rabinowitz, M. (1982) Proc. Natl. Acad. Sci. USA 79, 5847–5851.
[25] Mahdavi, V., Periasamy, M. and Nadal-Ginard, B. (1982) Nature 297, 659–664.
[26] Amara, S.G., Jonas, V., Rosenfeld, M.G., Ong, E.S. and Evans, R.M. (1982) Nature 298, 240–244.
[27] Grosskopf, R. and Feldmann, H. (1981) Curr. Gen. 4, 151–158.